DATA ANALYSIS ON THE BIKE SHARING DATASET

The link to the dataset and the Analysis Codes:
https://github.com/TheCreativeLad/Bike-Rental-Analysis

The Summary Statistics of the data

```
   instant         dteday         season  yr        mnth      holiday weekday workingday weathersit
Min.   :  1.0   Length:731      1:181   0:365   1     : 62   0:710   0:105   0:231     1:463
1st Qu.:183.5   Class :character 2:184  1:366   3     : 62   1: 21   1:105   1:500     2:247
Median :366.0   Mode  :character 3:188          5     : 62           2:104             3: 21
Mean   :366.0                    4:178          7     : 62           3:104
3rd Qu.:548.5                                   8     : 62           4:104
Max.   :731.0                                   10    : 62           5:104
                                                (Other):359          6:105
     temp             atemp            hum            windspeed          casual          registered          cnt
Min.   :0.05913  Min.   :0.07907  Min.   :0.0000  Min.   :0.02239  Min.   :   2.0  Min.   :  20   Min.   :  22
1st Qu.:0.33708  1st Qu.:0.33784  1st Qu.:0.5200  1st Qu.:0.13495  1st Qu.: 315.5  1st Qu.:2497   1st Qu.:3152
Median :0.49833  Median :0.48673  Median :0.6267  Median :0.18097  Median : 713.0  Median :3662   Median :4548
Mean   :0.49538  Mean   :0.47435  Mean   :0.6279  Mean   :0.19049  Mean   : 848.2  Mean   :3656   Mean   :4504
3rd Qu.:0.65542  3rd Qu.:0.60860  3rd Qu.:0.7302  3rd Qu.:0.23321  3rd Qu.:1096.0  3rd Qu.:4776   3rd Qu.:5956
Max.   :0.86167  Max.   :0.84090  Max.   :0.9725  Max.   :0.50746  Max.   :3410.0  Max.   :6946   Max.   :8714
```

From here we can see that for the casual customers (Customers who are not registered and do not use the system often):
- The maximum number of casual to use the system in a day is 1,410
- The minimum number of casual to use the system in a day is 2
- On average, the number of casual using the system in a day is 848

From here we can also see that for the registered customers (Customers who are registered and do use the system often):
- The maximum number of registered Customers to use the system in a day is 6,946
- The minimum number of registered Customers to use the system in a day is 20
- On average, the number of registered Customers using the system in a day is 3,656

For the total user (Both Casual and Registered):
- The maximum number of Customers to use the system in a day is 8,714
- The minimum number of registered Customers to use the system in a day is 22
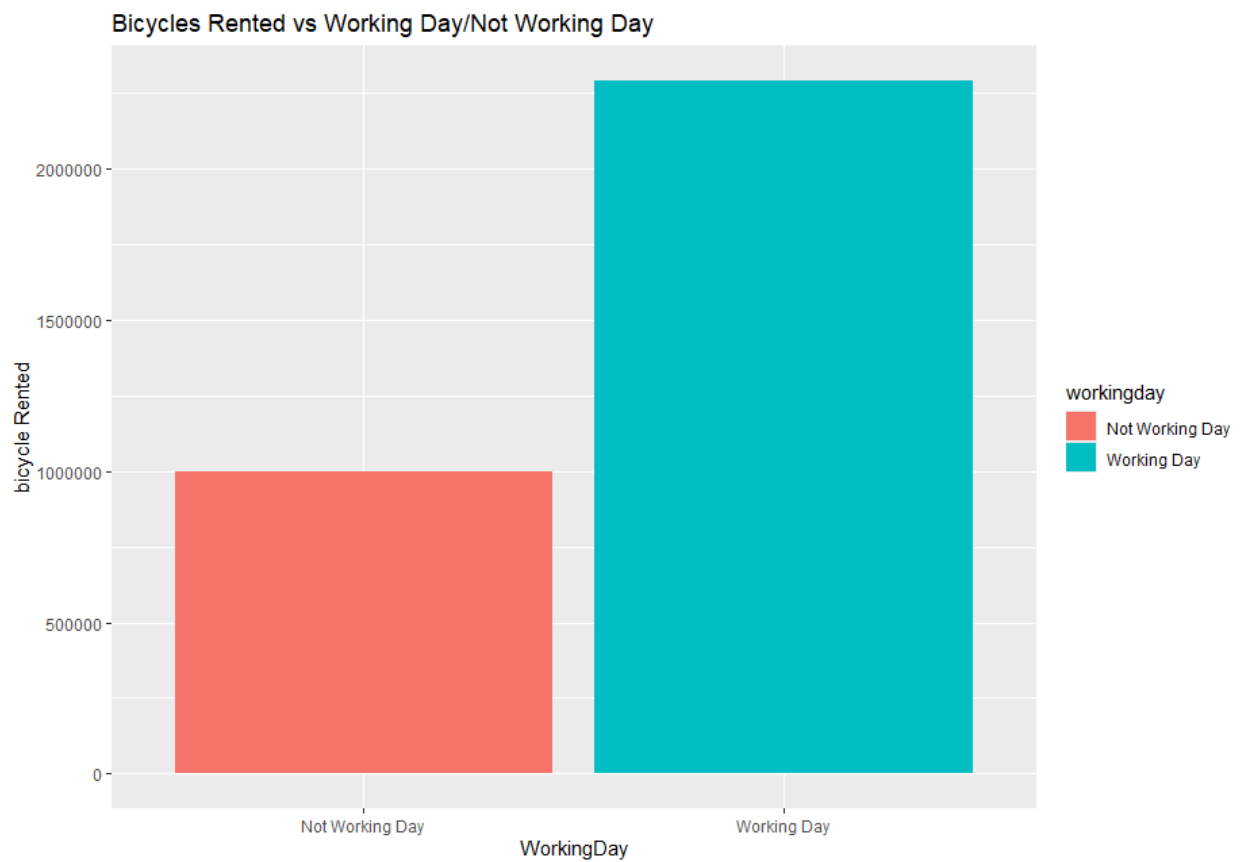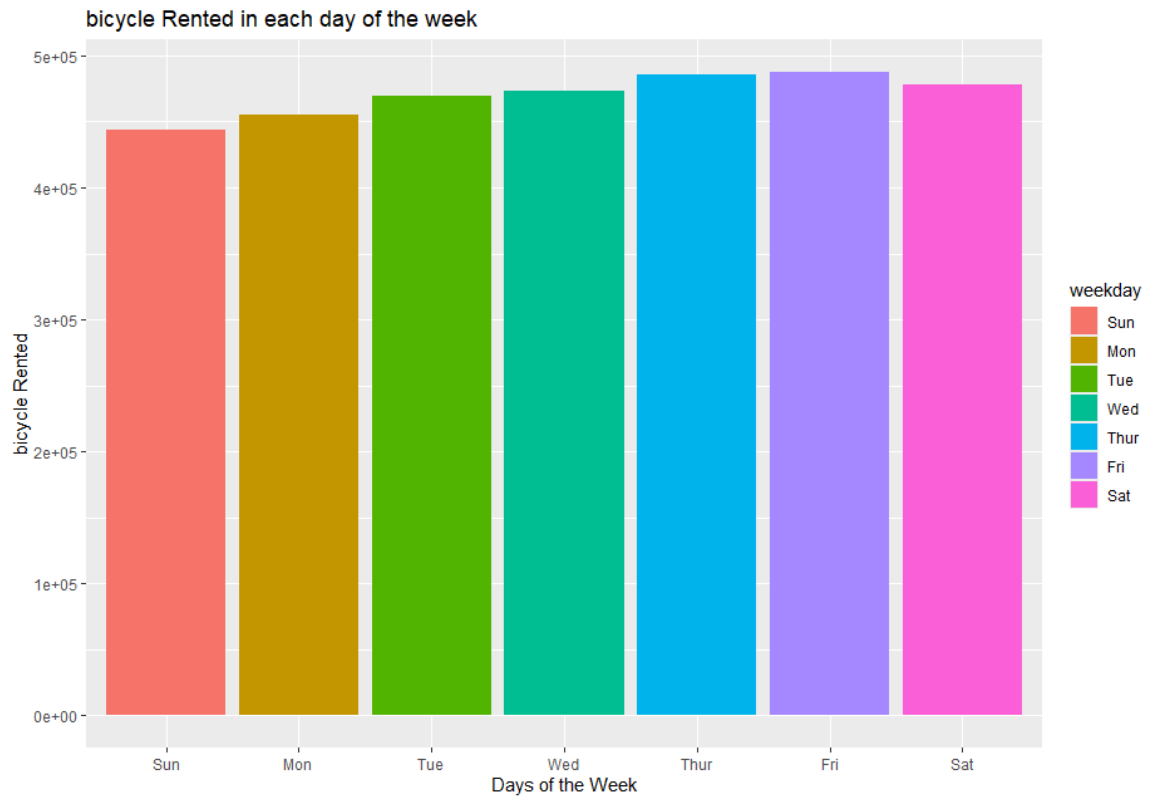- On average, the number of registered Customers using the system in a day is 4,504

CHECKING AND ANALYZING HOW DIFFERENT FACTORS AFFECTS THE RATE AT WHICH PEOPLE RENT BICYCLES

❖ **Checking how Public Holidays affects the Renting of Bicycles**

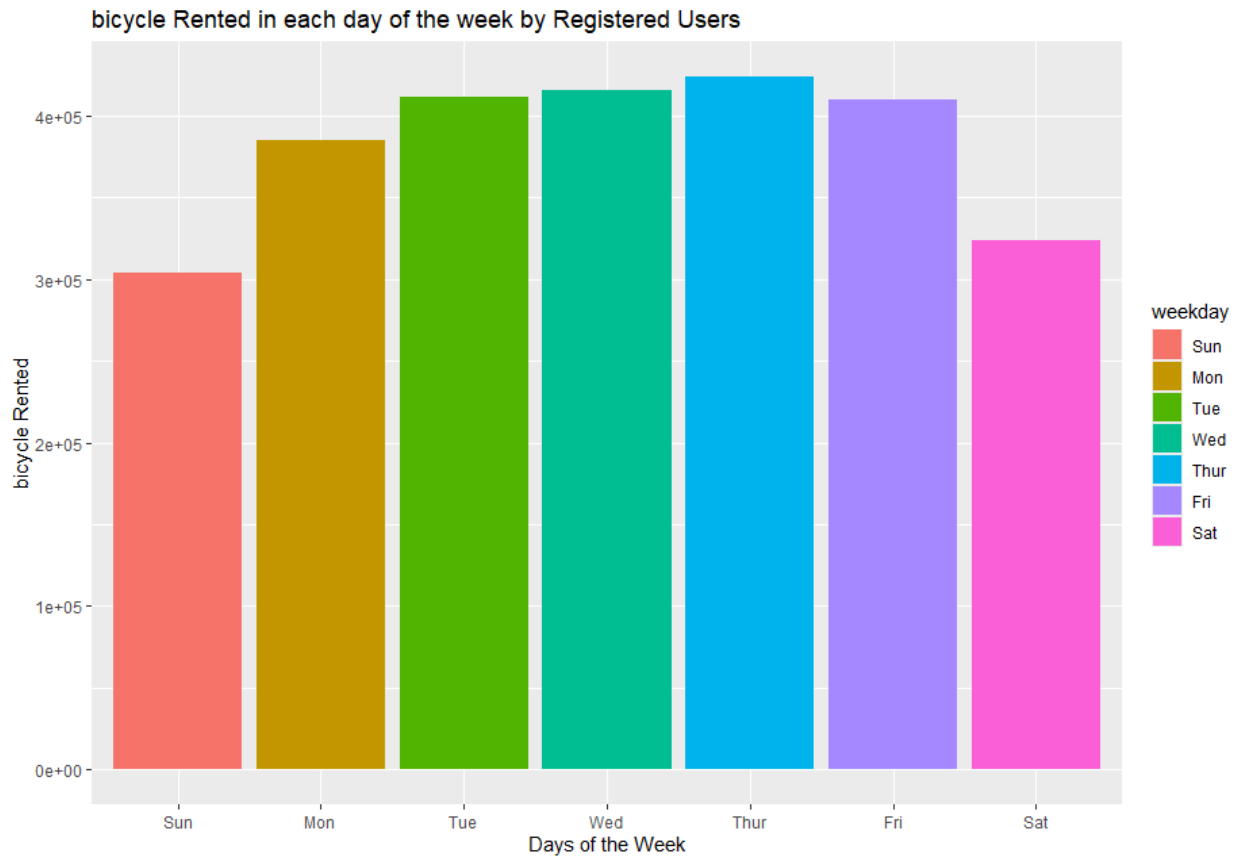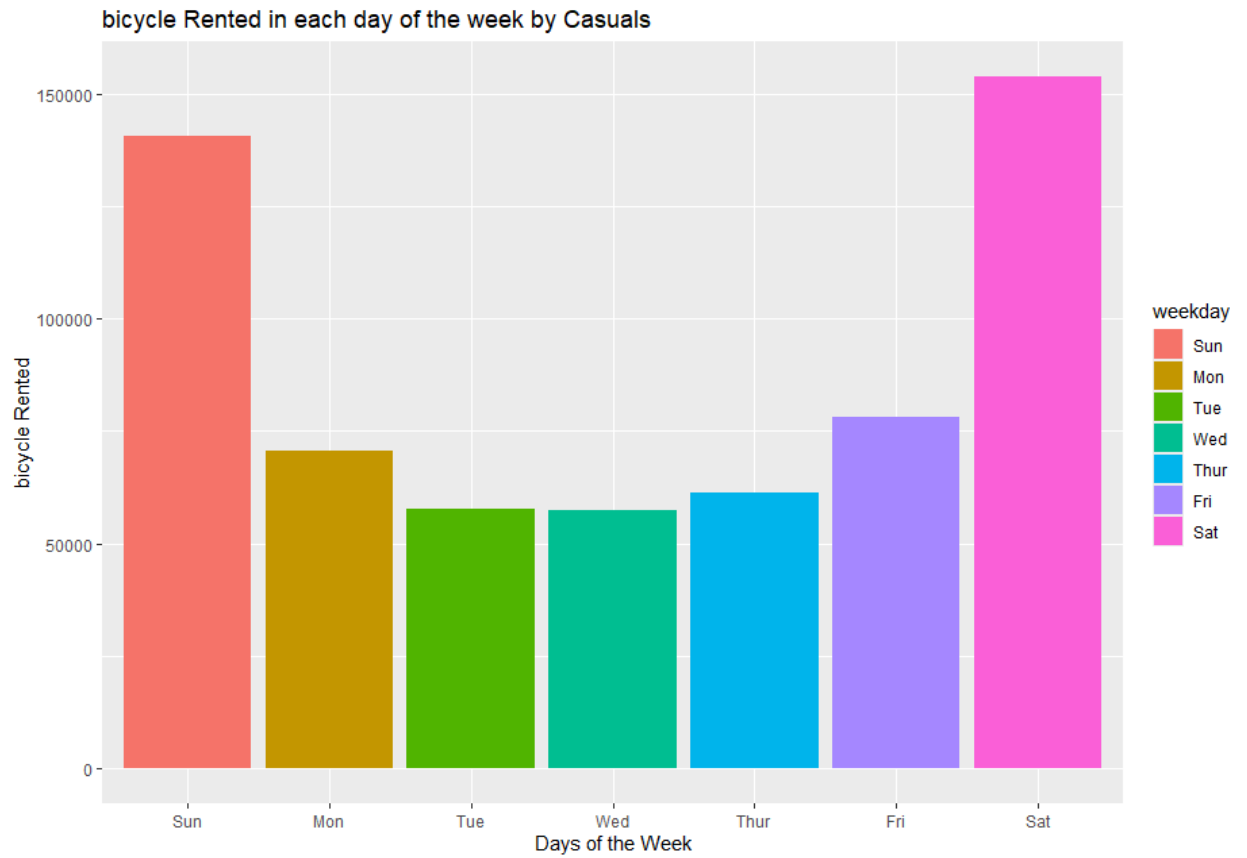## Relationship between Holiday and bicycle rented
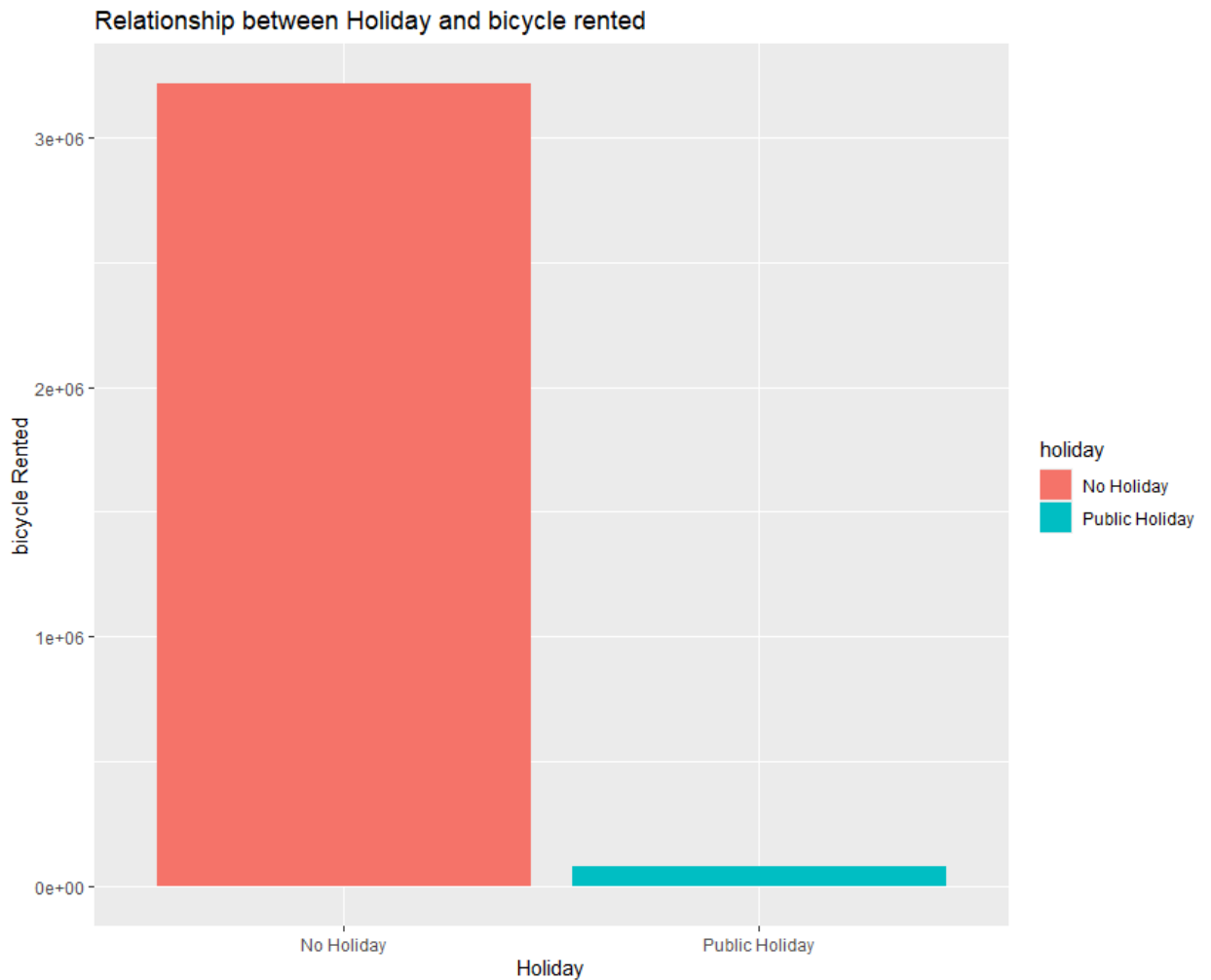
It is obvious the rate at which people borrow bikes on a public holiday is very small compared to if there is no public holiday. Probably because many people tends to stay at home on public holidays since there will be no going to work and sometimes, there might be restrictions in movement.

❖ **Checking how Public Holidays affects the Renting of Bicycles**

bicycle Rented in each day of the week

weekday
- Sun
- Mon
- Tue
- Wed
- Thur
- Fri
- Sat

Bicycles Rented vs Working Day/Not Working Day

workingday
- Not Working Day
- Working Day

We can see that people tend to rent bikes more during the working days of the week (Mon - Fri). And from the first plot, we can see that even during the week, people tends to rent bike more towards the end of the week, the movement is rising day by day from monday to friday, this may be due to stress, many people would have been stressed out towards the end of the week, and may decide not to trek.

bicycle Rented in each day of the week by Registered Users

bicycle Rented in each day of the week by Casuals

From the above charts, we can see that the casual user tends to use the bike more on weekends than working days while the registered Customers tend to use the bike more on working days and less on weekends. This may be that the casual Customers of ten rent the bike on weekends just for exercise and workout, so they do not register because they don't use it often but the registered Customers use it for getting themselves to work, so they register because they need it often and they won't want any delay while going to work during working days.

❖ **Checking how Weather Temperature affects the Renting of Bicycles**

The effect of Weather Temperature on the Bicycles Rented

$y = 950 + 7500\,x$
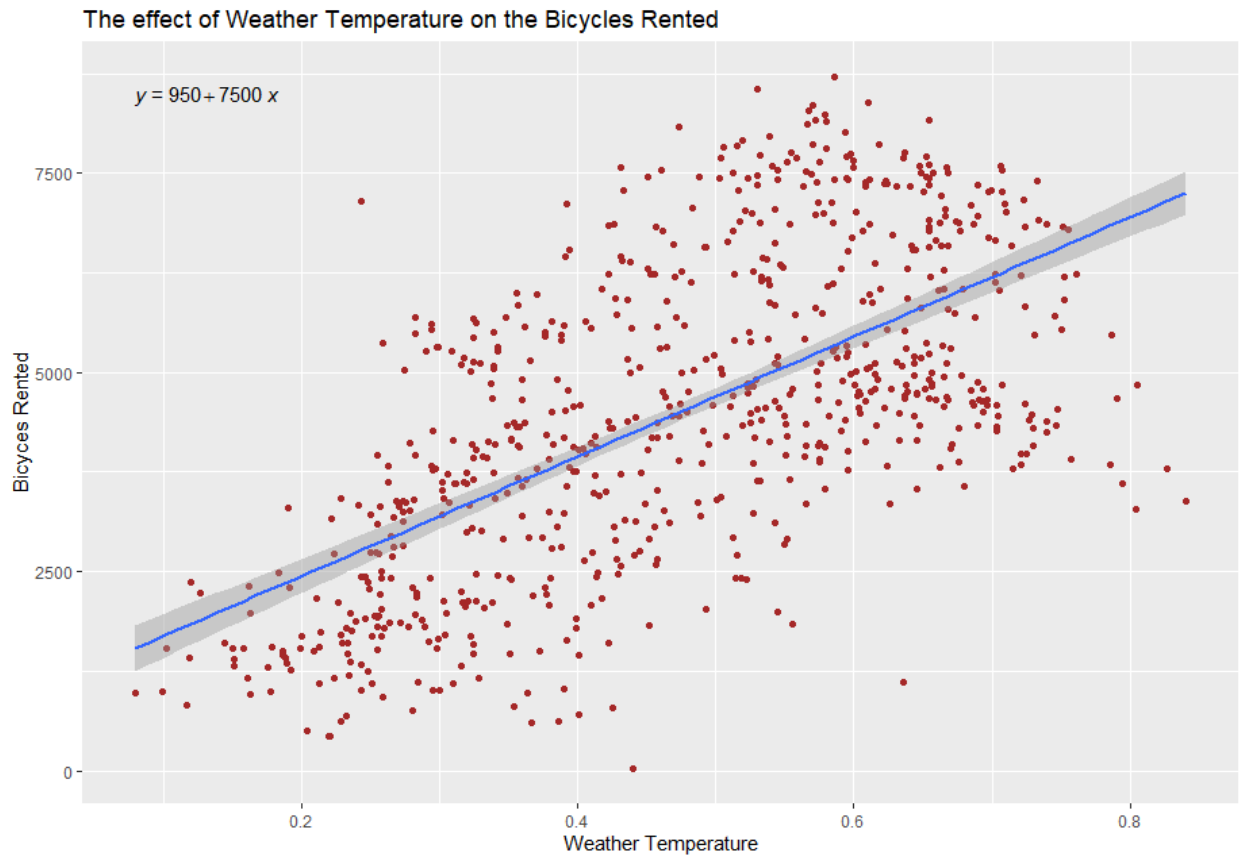
From the above plot, we can see that as the weather temperature increases, the number of bicycles rented also increases. Nobody likes to walk under a hot sun, so people tend to rent the bicycle more on a sunny day than a day where there is cool weather and the sun is not hot.

➢ **Checking the relationship between the Weather Temperature and the Bicycles rented using Linear Model**

```
Call:
lm(formula = cnt ~ atemp)

Residuals:
    Min      1Q  Median      3Q     Max
-4598.7 -1091.6   -91.8  1072.0  4383.7

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    945.8      171.3   5.522 4.67e-08 ***
atemp         7501.8      341.5  21.965  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1504 on 729 degrees of freedom
Multiple R-squared:  0.3982,    Adjusted R-squared:  0.3974
F-statistic: 482.5 on 1 and 729 DF,  p-value: < 2.2e-16
```

From the above result, we can see the linear model which is: y = 945.8 + 7501.8x, where y is the Bicycles rented and x is the weather temperature. This shows a positive linear relationship between them, and we see that the weather temperature is significant at all levels in predicting the number of bicycles rented. But we can only guarantee this significance by 40%, but we can go further by performing the correlation test to check how well they both correlated together.

```
        Pearson's product-moment correlation

data:  atemp and cnt
t = 21.965, df = 729, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5853376 0.6727918
sample estimates:
      cor
0.6310657
```

From the above result, we can see that the correlation between the weather temperature and the Bicycles rented is 0.63, which means the weather temperature can predict 63% of the Bicycles rented.

❖ **Checking how Humidity affects the Renting of Bicycles**

The effect of Humidity on the Bicycles Rented

$y = 5400 - 1400\,x$
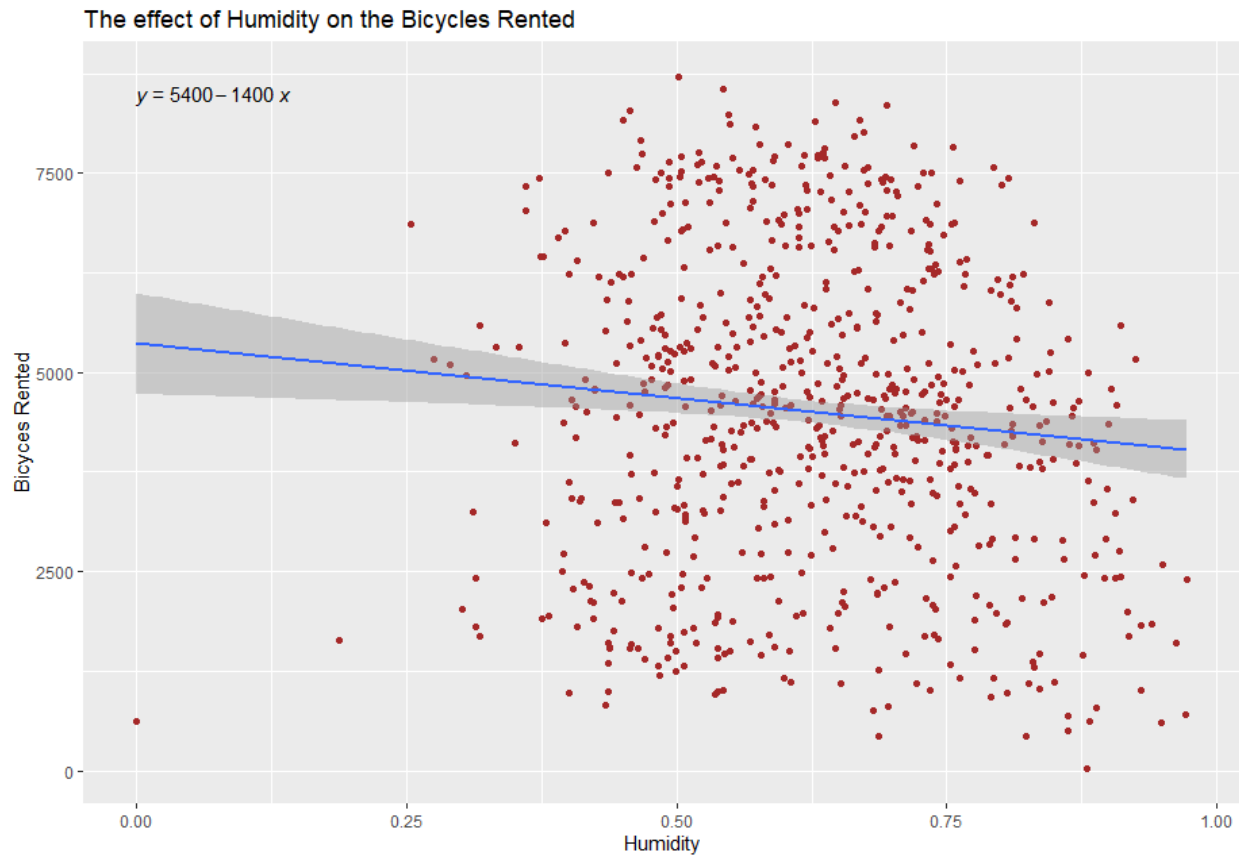
From the above chart, we can see that there is a relationship between humidity and the number of Bicycles rented per day. As the humidity increases, there is a decrease in the number of bicycles rented, this shows that many people prefer to walk or take other means to get to where they are going under a more humid weather condition. The weather is cool and taking a walk wouldn't be that much stressful on a day like this. But lets check further their relationship using Linear Model.

➢ **Checking the relationship between the Weather Temperature and the Bicycles rented using Linear Model**

```
Call:
lm(formula = cnt ~ hum)

Residuals:
    Min      1Q  Median      3Q     Max
-4741.0 -1386.9    50.3  1439.3  4036.8

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5364.0      322.7  16.623  < 2e-16 ***
hum          -1369.1      501.2  -2.732  0.00645 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1929 on 729 degrees of freedom
Multiple R-squared:  0.01013,   Adjusted R-squared:  0.008774
F-statistic: 7.462 on 1 and 729 DF,  p-value: 0.006454
```

From the above result, we can see that the linear model is: y = 5364.0 - 1369x where y is the bicycles rented and x is the humidity. We can also see that humidity is significant in predicting the number of bicycles rented. Performing the correlation test to check how well they correlated with each other.

```
        Pearson's product-moment correlation

data:  hum and cnt
t = -2.7317, df = 729, p-value = 0.006454
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.17191732 -0.02835191
sample estimates:
       cor
-0.1006586
```

We can see that there is a negative correlation of -0.1, which means that there is a negative relationship of 10% between humidity and the number of bicycles rented.

❖ **Checking how Different Seasons affects the Renting of Bicycles**

Total Bicycles rented in each Season
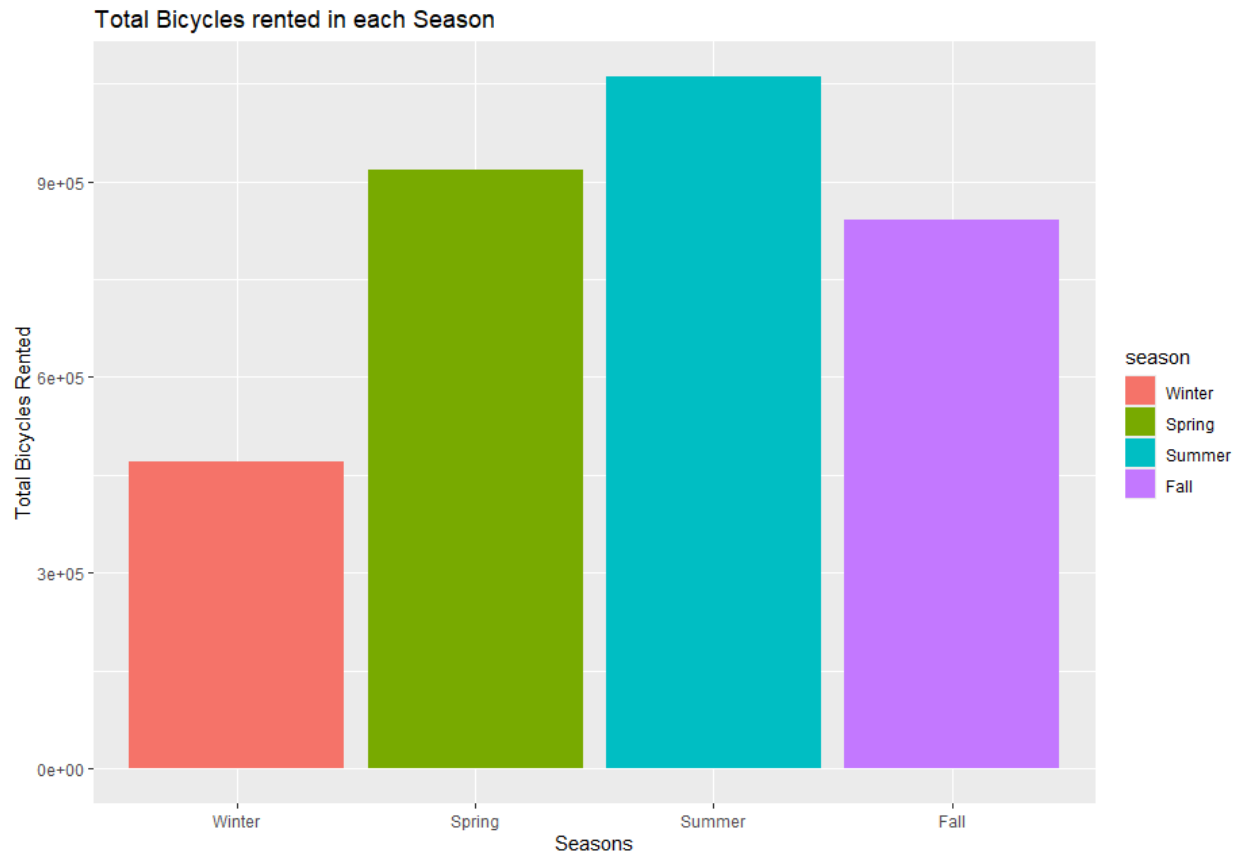
From the chart above, we can see that the rate at which people rent bicycles at every season differs. From this data now, we can see that people rent bicycles more during the Summer season, then during Spring, Fall and winter in that order. During the Winter season, there is always heavy snow, ice accumulation, freezing temperatures and wind chill which is not a good condition for riding a bicycle, so we can see why there is a decrease in the number of bicycles rented. Also during the Fall, there is much rain which is a factor that can affect the number of bicycles rented in this season.

❖ **Checking how Different Seasons affects the Renting of Bicycles**

## Effect of the weather condition on the Bicycles rented
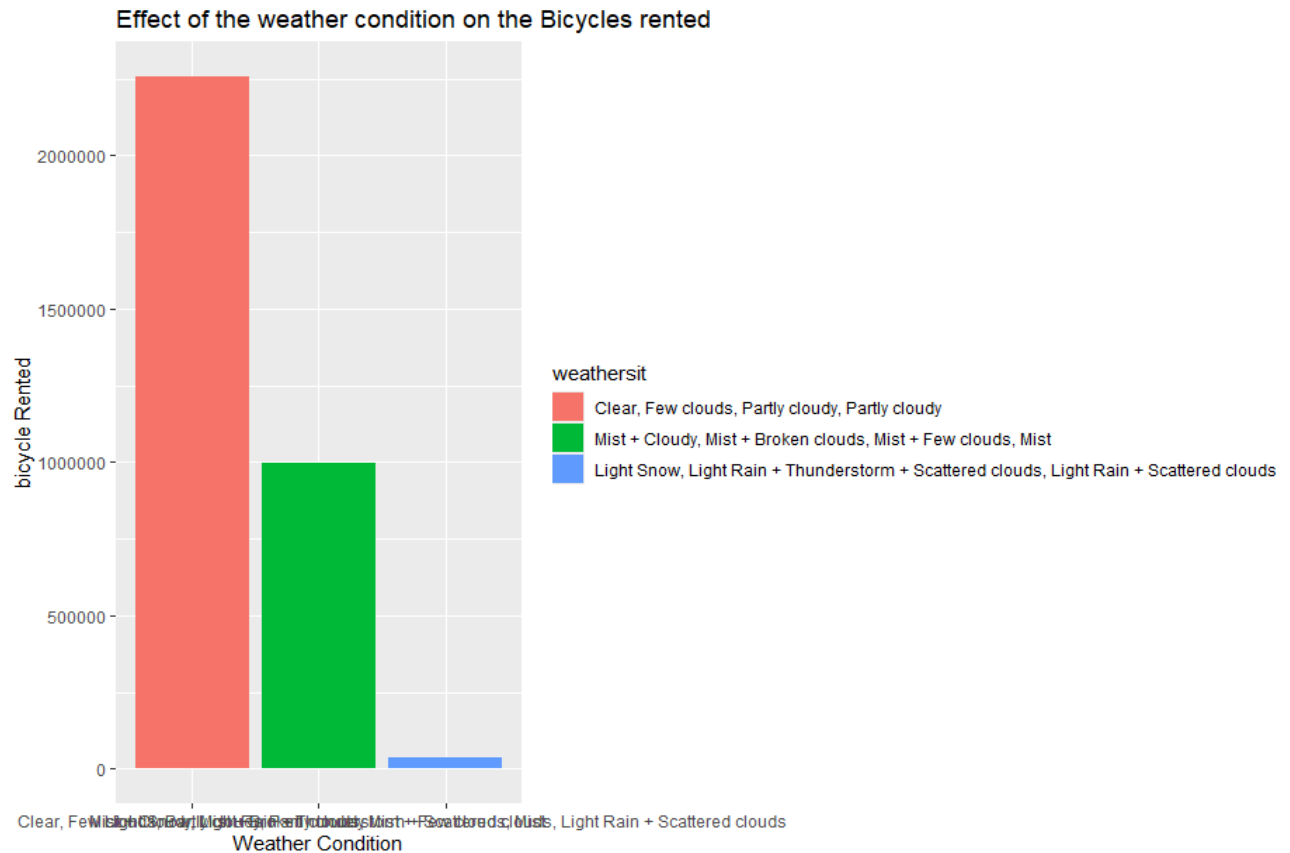
From the above chart, we can see that there is a huge impact of weather conditions on the number of bicycles rented in a day. People tend to rent more bicycles on a day when the weather is clear with few clouds or partly cloudy than on a day where the weather is misty with cloudy or snowy weather.

❖ **Checking the Renting of Bicycles across each Months**

Total Bicycles rented in each Month
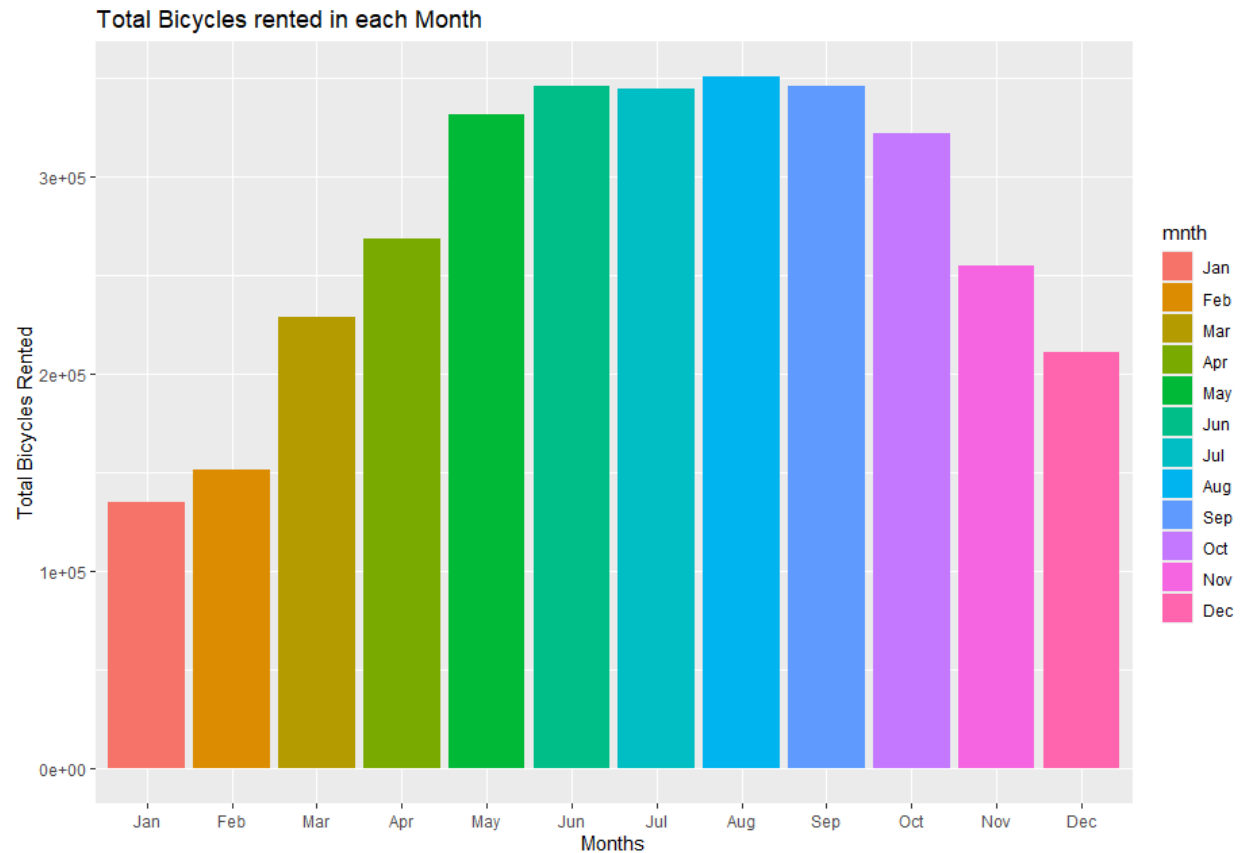
From the above chart, we can see that people tend to rent more bikes during the months of May to October, which can be related to the seasons. We can see that the month of December to February is the winter season, and as seen above earlier, people tend not to rent bicycles during the winter season.

❖ **Checking how the Wind Speed affects the Renting of Bicycles**
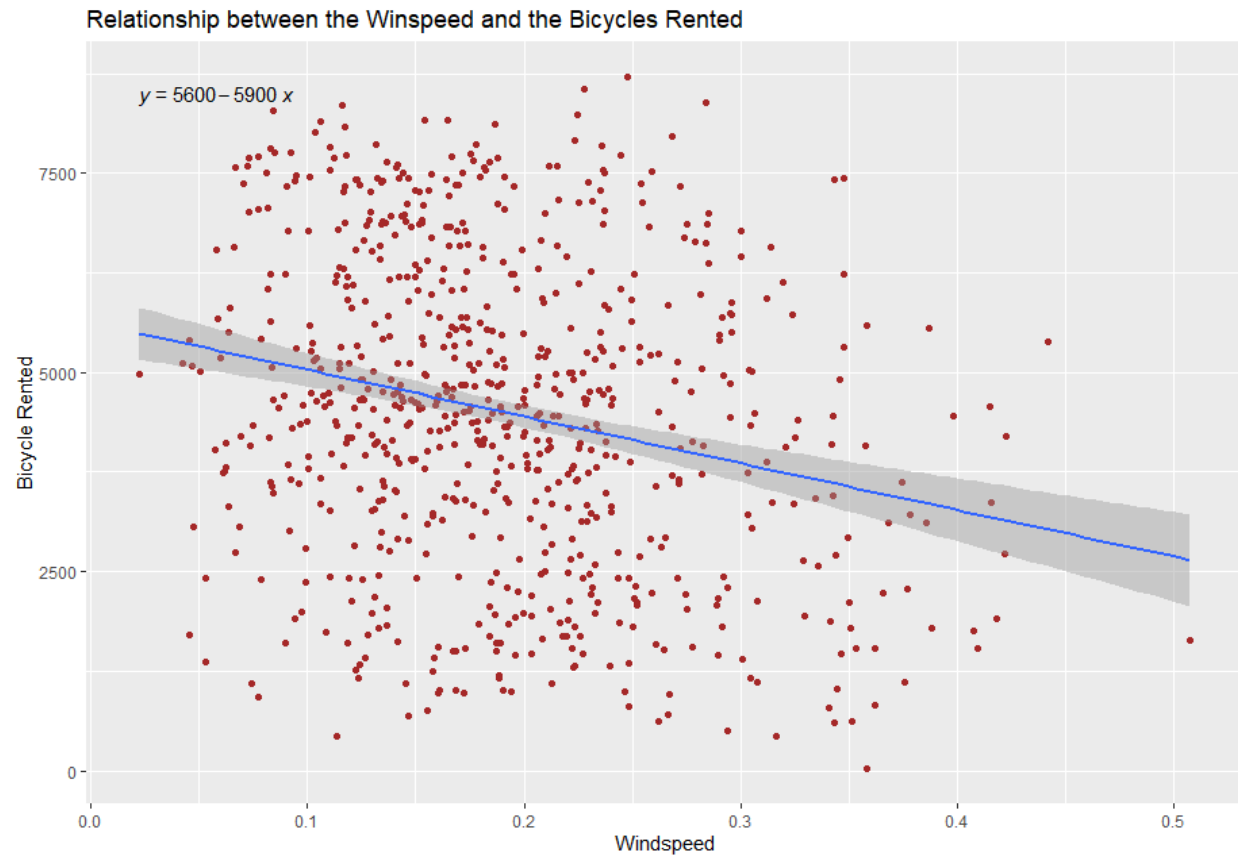
Relationship between the Winspeed and the Bicycles Rented

$y = 5600 - 5900\,x$

From the chart above, we can see that there is a relationship between the Wind Speed and the number of bicycles rented on each day. We can see that as the wind speed is increasing, the number of bicycles rented is decreasing. So people tend not to rent bIcycles on a day when the wind speed is high.

➢ **Checking the relationship between the Wind Speed and the Bicycles rented using Linear Model**

```
Call:
lm(formula = cnt ~ windspeed)

Residuals:
    Min      1Q  Median      3Q     Max
-4522.7 -1374.7   -74.6  1461.8  4544.0

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5621.2      185.1  30.374  < 2e-16 ***
windspeed    -5862.9      900.0  -6.514 1.36e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1884 on 729 degrees of freedom
Multiple R-squared:  0.05501,   Adjusted R-squared:  0.05372
F-statistic: 42.44 on 1 and 729 DF,  p-value: 1.36e-10
```

From the above result, we can see that the linear model is: y = 5621.2 - 5862.9x where y is the bicycles rented and x is the Wind Speed. We can also see that Wind Speed is significant in predicting the number of bicycles rented. Performing the correlation test to check how well they correlated with each other.

```
        Pearson's product-moment correlation

data:  windspeed and cnt
t = -6.5144, df = 729, p-value = 1.36e-10
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.3019236 -0.1648349
sample estimates:
      cor
-0.234545
```

We can see that there is a negative correlation of -0.23, which means that there is a negative relationship of 23% between Wind Speed and the number of bicycles rented.

❖ **Checking the ratio of the Casual User and the Registered CustomersRenting of Bicycles**

From the analysis, we can see that there are about 19% of the Casual Customers and about 81% of the registered customers